DOCUMENT RESUME

ED 338 641 TM 015 967

AUTHOR Corbett, H. Dickson; Wilson, Bruce

TITLE Raising the Stakes in Statewide Mandatory Minimum Competency Testing.

INSTITUTION Research for Better Schools, Inc., Philadelphia, Pa.
SPONS AGENCY Office of Educational Research and Improvement (ED), Washington, DC.
PUB DATE 89
NOTE 25p.
AVAILABLE FROM Research for Better Schools, 444 North Third Street, Philadelphia, PA 19123.
PUB TYPE Reports - Research/Technical (143)

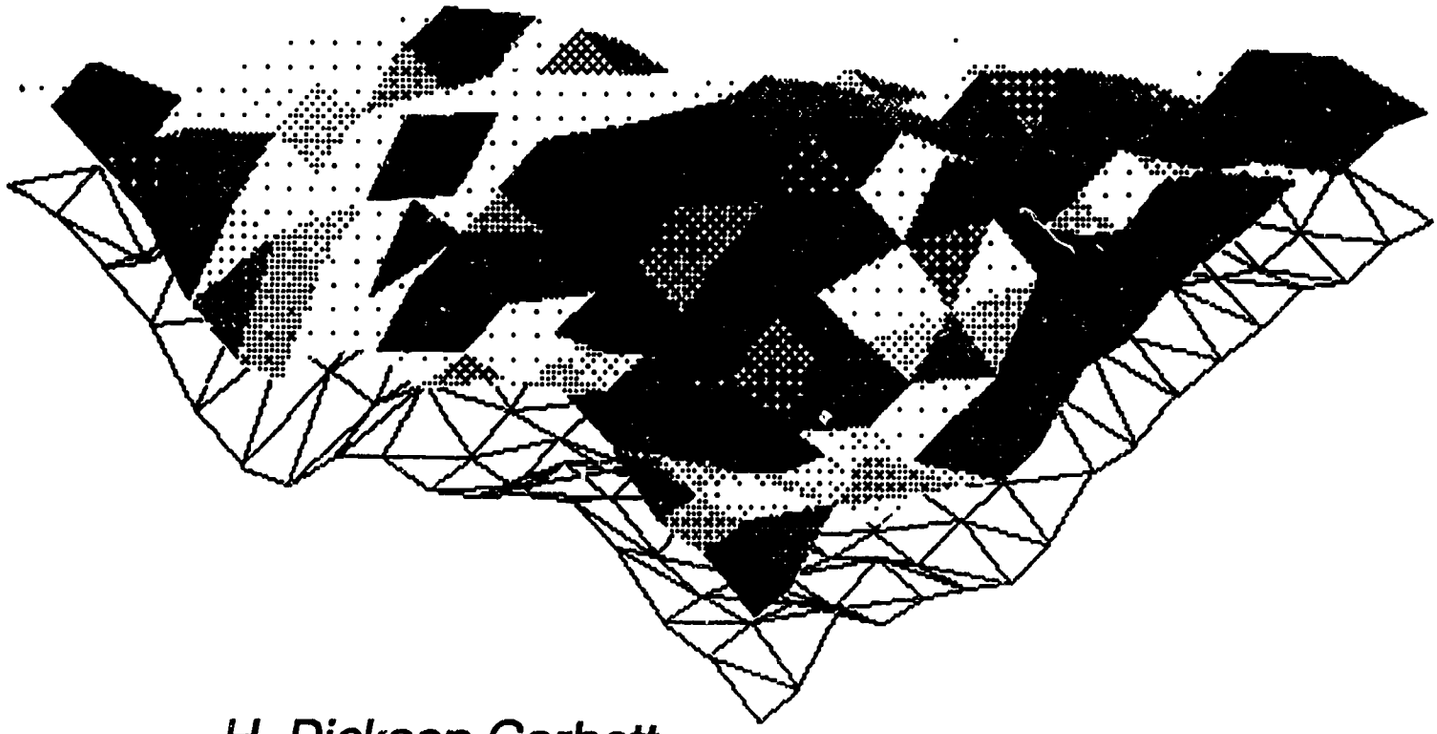
EDRS PRICE MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS Educational Assessment; *Educational Change; Elementary School Teachers; Elementary Secondary Education; Graduation Requirements; *Minimum Competency Testing; Principals; Questionnaires; School Personnel; *School Surveys; Secondary School Teachers; *State Programs; Testing Problems; *Testing Programs; Test Results
IDENTIFIERS *High Stakes Testing; Maryland; Pennsylvania; Stakeholder Survey

ABSTRACT

The mandated minimum competency programs of two states--Pennsylvania and Maryland--are examined, and some of the effects on school districts of raising the testing stakes are reviewed. In a survey conducted during the winter of 1986-87 in Pennsylvania and Maryland, one teacher, one principal, and one central office staff member each from 277 Pennsylvania districts and 23 Maryland school systems replied to a questionnaire on the testing program. The stakes increased in Maryland due to the approach of the time when Maryland students would be responsible for passing all four state competency tests to graduate. The stakes increased in Pennsylvania due to a brief public release of school district rankings based on test scores from the spring of 1987 test administration. The survey results indicate that school districts in higher stakes testing situations make more adjustments to instruction and organization than do districts in lower stakes situations. In both states, the perception of higher stakes associated with testing resulted in an intensification of the pressure on local educators to improve test scores, which in turn stimulated changes in local practices. High stakes statewide testing programs were seen to alter the political character of districts by increasing the probability that community elements could and would exercise influence. Such effects of high stakes testing could counteract efforts to reform teaching. (SLD)

TM

Raising the Stakes in Statewide Mandatory Minimum Competency Testing



H. Dickson Corbett
Bruce Wilson

RBS
Research for Better Schools
444 North Third Street
Philadelphia, PA 19123

Raising the Stakes in Statewide Mandatory Minimum Competency Testing

H. Dickson Corbett
Bruce Wilson

RBS
Research for Better Schools
444 North Third Street
Philadelphia, PA 19123

This publication is a product of RBS' Applied Research Project, Bruce Wilson and H. Dickson Corbett, Directors and the R&D Project, Keith M. Kershner, Director.

Word Processing by Arlene Large and Carol Crociante
Graphic Art by Peter Robinson

This publication is based on work sponsored, in part, by the Office of Educational Research and Improvement (OERI), U.S. Department of Education; the content of this publication does not necessarily reflect the views of OERI, the Department, or any other agency of the U.S. Government. Portions of this work were made possible through the cooperation of the Maryland State Department of Education (MSDE). The opinions expressed, however, do not necessarily reflect those of MSDE either.

Printed in the United States of America.

Table of Contents

INTRODUCTION

One manifestation of educational reform in this decade has been the use of statewide, mandatory, high-stakes tests -- particularly in certifying professionals and encouraging student attainment of certain minimum competencies. The level of the stakes associated with a test is the extent to which test performance is perceived by administrators, teachers, students, and/or parents to be "used to make important decisions that immediately and directly affect them" (Madaus, 1988:86). In the case of minimum competency testing (MCT) -- the type of statewide testing with which this publication is concerned -- connecting test results to student promotion or graduation raises the stakes associated with the test and increases the seriousness with which educators and citizens regard the state's program. Whether the ensuing activity at the local level reforms systems for the better remains unanswered, and consequently so does the advisability of a state's use of higher stakes as a policy lever to instigate that activity.

This publication looks specifically at two states' mandated MCT programs, discusses some of the effects on school districts associated with raising the testing stakes, and makes several recommendations regarding a state's use of stakes. The argument is that as the stakes of statewide MCT rise, the testing program is indeed taken more seriously at the local level, especially in terms of matching local objectives to those covered in the test and in terms of resequencing course content to insure that content contained on a test is covered in classrooms prior to the test. However, at some point during an increase in stakes, pressure on a district can intensify such that a shift in local focus occurs, and student performance on the test becomes an end in itself rather than merely an indicator of student attainment of broader learning outcomes. The consequence is that educators in the district begin to question whether their efforts to improve specific test scores are consistent with their interest in promoting student learning. The policy challenge is to encourage local attention to reform without instigating counterproductive responses.

STAKES AND TESTING

The literature on the effects of various changes in state educational MCT testing policies is scant (Madaus, 1988; Stake et. al, 1987), but six general investigations of high-stakes testing provide at least a starting point for examining the topic. Relying heavily on anecdotes, testimony from public hearings, historical accounts, and an occasional international study, Madaus (1988:88-98) induced seven principles regarding the relationship between the level of stakes a test is perceived to have and the effects of the test on action at the local level:

- The power of tests and examinations to affect individuals, institutions, curriculum or instruction is a perceptual phenomenon: if students, teachers, or administrators believe that the results of an examination are important, it matters very little whether this is really true or false -- the effect is produced by what individuals perceive to be the case.

- The more any quantitative social indicator is used for social decisionmaking, the more likely it will be to distort and corrupt the social processes it is intended to monitor.

- If important decisions are presumed to be related to test results, then teachers will teach to the test.

- In every setting where a high-stakes test operates, a tradition of past exams develops, which eventually de facto defines the curriculum.

- Teachers pay particular attention to the form of questions on a high-stakes test, (e.g. short answer, essay, multiple choice), and adjust their instruction accordingly.

- When test results are the sole or even partial arbiter of future educational or life choices, society tends to treat test results as the major goal of schooling rather than as a useful but fallible indicator of achievement.

- A high-stakes test transfers control over the curriculum to the agency which sets or controls the exam.

This list emphasizes that stakes can become high when test results automatically trigger important consequences for students or the school system, and also when educators, students, or the public perceive that significant consequences accompany test results. Thus, a formal trigger of consequences need not be built into the testing program for stakes to be high. Instead, test results can cause the public to make an assessment of the quality of the school system that serves them, and this judgment in turn can lead to a conclusion that children's choices of post-secondary schooling or occupation have been affected. The product of this process is increased public pressure to improve test scores when the perception is that the system is likely to have a negative impact on those choices. Such was the case in Kentucky (Center for the Study of Testing, Evaluation, and Educational Policy, 1986) and such was the case in one state discussed later.

Murnane (1987) identified three common district responses to high-stakes conditions: excluding low-scoring children on some basis from taking the test,

focusing instruction on the skills measured on the tests, and teaching test-taking skills. He notes, however, that:

> *...publicizing outcome data for individual schools and school districts may be a relatively effective strategy by which states and the federal government can persuade local school districts to concentrate on improving student learning. On the other hand ...the responses of local school officials could result in improved average test scores without increasing student learning. In this case the publicized test scores provide misleading information and the responses by local officials reduce the effectiveness of the organizations that they lead (Murnane, 1987:105).*

Thus, Murnane, like Madaus, argued that there is the potential for distorted, counterproductive local behavior under high stakes conditions.

Three empirical studies of district high-stakes testing programs also note the potential for similar effects. Polemini (1977) found test security in a large city's testing program to be a problem as local educators sought to obtain advanced copies of the test, primarily because they feared job accountability would be tied to results. First and Cardenas (1986) claimed that districts excluded certain categories of students, particularly those who would likely do poorly, from the test-taking pool as a way to boost test results. LeMahieu (1984) discovered that local, high-stakes tests could be beneficial, but great care had to be taken to avoid having staff make testing objectives the sole content covered in classes.

It would seem that high-stakes tests are taken seriously, if not always productively, at the local level -- at least in terms of local staff perceptions that they have to address test results. Increasing the stakes, then, is a means of increasing the pressure on local systems to alter their operation. From the state perspective, such pressure may be a critical ingredient in promoting successful improvement at the local level, according to findings from a ten-state study of state-initiated school improvement reported by Anderson et. al (1987). The same researchers also state that "more important than the type of pressure was the fact that it existed" (Anderson et. al, 1987:74). This paper argues that the type of pressure does matter: Pressure via raised stakes encourages local action, but this action may be contradictory to the intended goals of reform.

The next two sections discuss the effect of the level of MCT stakes on local action, first in terms of the seriousness with which districts regard the tests, and then in terms of a shift in district focus from long-term learning objectives to short-term test score improvement.

STAKES AND HOW SERIOUSLY THE PROGRAM IS TAKEN AT THE LOCAL LEVEL

An important estimate of the seriousness with which a program is taken is the extent to which local activity is adjusted in response to the test. Results of a survey Research for Better Schools (RBS) conducted during the winter of 1986-87 in Pennsylvania and Maryland bear out the expectation that school districts in higher stakes testing situations make more adjustments in instruction and organization than those in lower stakes situations. A questionnaire that solicited information concerning the administration of the testing program, test uses, test impacts, and school system context was completed by a teacher, principal, and central office staff member in 277 of Pennsylvania's 501 districts and by three occupants of each position in 23 of Maryland's 24 systems. Below is a brief summary of the conclusions. (See Corbett and Wilson, 1987 for a complete discussion of the study.)

The two states designed their testing programs such that there were at least four important differences. First, in Pennsylvania, the purpose was to identify students who were failing so that they could receive remediation determined by the district. Students were not required to retake the test to achieve a passing score. In contrast, Maryland made a passing score on four separate tests a prerequisite for graduation. At the time this paper was written, the first cohort of students required to pass all four were juniors. Thus, one year remained before the testing program reached its most stringent point. Special education students who did not meet this requirement could receive a certificate of attendance. Second, Pennsylvania students took their tests in the third, fifth, and eighth grades. Maryland tested students beginning in ninth grade, although a practice test was administered in the middle school. Third, the legislature in Pennsylvania approved a special appropriation to assist local remediation, whereas Maryland offered no financial assistance for this purpose. Fourth, Pennsylvania's test was a legislative response to the calls for educational reform issued by the commissions and panels convened in the early 1980's. Although educators in the state suggested test objectives, commercial test publishers were invited to bid on a contract to provide the state's instrument. Maryland initiated a statewide curriculum improvement program several years prior to beginning the testing program, with the expressed purpose of anticipating the instructional quality necessary to perform well on the tests. Moreover, educators around the state were selected by the SEA to provide input into the content and form of the tests.

Clearly, Maryland's program should have had a greater impact on its local systems than Pennsylvania's program, primarily because Maryland's policy insinuated itself into an important organizational event -- graduation -- and because preceding statewide improvement and actual test development activities engendered a cumulative anticipation of the day the tests would be put into place.

According to the RBS survey, this proved to be the case. Essentially, in comparison with Pennsylvania, Maryland school systems focused more directly on improving their test scores, altered their curriculum to a greater extent (especially in terms of redefining course objectives and resequencing course content), and more often used the scores to compare school performances both within the district and across school systems. Maryland educators also reported that students tended to take school more seriously, and those with special learning needs were better known and received more attention. At the same time, Maryland teachers were reported to be under greater stress, to have more paperwork, and to have experienced decreased reliance on their professional judgment than teachers in Pennsylvania. Regarding these last findings, interviews with Maryland educators subsequent to the survey revealed that these changes in teachers' work lives were largely concomitants of self-induced pressure to make sure their students succeeded. That is, regardless of their personal and professional opinions about the tests, the fact was students had to pass them in order to graduate, and teachers felt responsible to ensure that their students did so.

In addition to information concerning curriculum adjustments, the survey asked respondents to assess whether the adjustments were for the better. The state-to-state differences were again dramatic and consistent. In Maryland there was a much stronger feeling that the state-mandated MCT program had narrowed and improved the curriculum in terms of both course objectives and the range of courses offered. Local educators explained that this assessment of the curriculum was the consequence of aligning the curriculum with test objectives. A clearer definition of what was expected to be covered represented an improvement over rambling curriculum guides, but at the same time did exclude some content that staff members previously had deemed worthy of inclusion. Up to a point, Maryland educators viewed a tighter curriculum as a better one; they worried, however, that the trend would lead to excessively basic course offerings.

Maryland educators also believed their systems had become more focused on testing than learning, and experienced a greater sense of discontinuity between the testing program and what they felt should be taught than did Pennsylvania educators. These latter two effects became exacerbated in the year following the survey. Those subsequent developments are the topic of the next section.

EFFECTS OF RAISING THE STAKES

The survey discussed above presented a snapshot of the differences in educators' reactions to two state-mandated testing programs. The picture was taken in the late fall of 1986 and the early winter of 1987. Events in both states

subsequent to the survey, however, had significant effects on educators' perceptions of the tests. In both states, the testing stakes increased -- due to a brief public release of school district rankings based on test scores in Pennsylvania and to the approach of the time when Maryland students would be responsible for passing all four of the tests to graduate, two of which Maryland educators reported were particularly troublesome. Field interviews RBS conducted in 11 school districts in the two states during the fall of 1987 as a followup to the survey elicited comments concerning the local effects of raising the stakes.

Pennsylvania

The key event in Pennsylvania was the publication of the results from the spring of 1987 test administration. Rather than the customary low-key sending of the scores to districts for each to handle as it saw fit, the event was orchestrated by the chief state school officer (CSSO). In a public media briefing, the CSSO provided documents that ranked school systems in the state in terms of the percentage of students who passed the cut-off point. A subpopulation of schools that had achieved a 100 percent passing rate despite a "high-risk" student population was singled out as being "poised on the brink of excellence," and other subgroups of "improving" schools were lauded. To cap off the presentation, the CSSO touted the tests as the best measure available to assess the effectiveness of Pennsylvania's schools. An immediate protest to this use of the scores arose from educators across the state and resulted in the withdrawal of the documents containing the rankings. This reaction was intelligible not only in terms of the conflict between the rankings and local views of the purposes of the testing program but also, as Fuhrman (1988) makes clear, in terms of the more subtle role the Pennsylvania SEA traditionally adopted in its interactions with districts.

The withdrawal of the rankings did not strike the event from either educators' or their communities' emotional record, however. Educators in three of the six Pennsylvania districts visited argued that the "game" had now changed in their systems:

> *The purpose of the test changed in September. It is no longer for remediation, but to rank order schools. [District 1 superintendent]*

> *The results should be between the state and the school district if the test is to help. When they release scores and say 58 kids need help, we can say we've already identified 40 of them. But the negativism starts; it starts [phone] calls and there is no question I now have pressure on me. [District 2 superintendent]*

The test was not all that important ...But we might as well face up to it; with the publication of school by school results ...one of the goals will be to raise the percentage above the cut score. [District 3 assistant superintendent]

Of the remaining three districts, one -- an urban system -- had "bought into" the test early in the program and had already begun using the scores comparatively. In fact, interview subjects in this system, to a person, pointed with pride to several of the schools that had achieved "high" passing rates relative to the student population they served. The scores were already highly visible in the community and the CSSO's actions contributed little additional publicity to how the schools were doing. In another district (which was rural), the community had taken little interest in the scores and, according to the superintendent, the system did not need to treat the test as other than a means of identifying students for additional instruction. In the third, an assistant superintendent claimed that "the publication of scores was deplorable; it was never the intent to rank schools." The person asserted that the scores would be downplayed in the district as they had been in the past.

What really seemed to be changing for the first set of three districts in Pennsylvania was the stakes; they got higher, as a result of the increased visibility of score comparisons and the subsequent increased, albeit reluctant, acceptance of the scores as a benchmark -- that is, as a widely recognized point of reference when discussing the performance of schools in the district and surrounding districts.

Staff in the three districts reported they did not believe the tests to be particularly important educationally, and they did not embrace the tests as valid indicators of attainment. They nevertheless acknowledged that they already were or would soon be treating the scores more seriously than in previous years. As one disgruntled educator claimed, what once was an educational tool had now become a weapon.

A central office administrator in District 3 commented,

The tests are not all that important. We use our own standardized testing program to modify instruction.

But since the publicity surrounding the scores had increased, more attention had been given to the tests. According to that administrator:

One thing we did was to say "here are the objectives on which the test was developed, look at them and see if they are being covered." This

didn't result in change but now that they [SEA] are publicizing the test scores more people who felt they could put the test aside will look at it and say not only have I covered it but do I feel the students will do well? Before I don't think there was as serious a reaction to analyze and interpret the schools' program as there probably is now.

Additional impetus for emphasizing test objectives in this same district came when a six percent difference in the number of students passing occurred between two middle schools in the system. Despite the fact that both had passing rates above 89 percent, the administrator went on to say:

We couldn't come up with an answer [for the difference] although the lower [scoring] school said they didn't think they needed to take it seriously. My response is you'd better. We might as well face up to it. One of the goals is going to be to raise the percentage of students above the cut score; so if you're not now emphasizing the test, you'd better. It may not be a legitimate impact, but it is there. The danger is not keeping it in proportion. We need to understand what the tests' place is and that's the danger in how the results are now being emphasized and publicized.

In District 1, a problem arose when surrounding districts' scores matched those of the system, even though the superintendent felt that its carefully and systematically developed curriculum far surpassed the offerings of those around them. The response?

We don't believe in the tests that strongly, but we will be forced to see all material is covered before the tests. We definitely are going to do it. We won't be caught in the newspapers again. [superintendent]

The brunt of not "getting caught" was placed on the reading program -- a recently revised, developmental curriculum. The timing of the test administration required shifting the sequence of topics to be covered. An outraged reading coordinator responded:

You have to alter a curriculum that is already working well and so [now] we can't follow the developmental process already established. Kids are already growing in a structured program; but it [pressure to change] comes from the board, community, and adverse publicity.

The superintendent empathized with the coordinator:

I don't have much faith in the tests. I don't want to change the curriculum, and it's not a major revision, but we've got to do better.

Still, it's not the right thing to do to anyone. I don't want to over-react, but I'm also going to have to spend time on things I shouldn't have to do: public relations, testing meetings -- just to make the board feel comfortable. It'll never happen again when we see a worse district doing better than us.

The actions were to be undertaken in a context similar to District 3, where standardized tests had long been an integral part of school improvement.

We feel you can't toy with nationwide standardized tests. That's what we believe in, and our performance has been very good. But over the next seven months, we'll be publishing more things about standardized tests and our interpretations of the [state] test scores.

District 2 administrators also indicated a preference not to alter a systematic process for addressing curriculum issues. The district took a cyclical approach, working on one content area at a time according to a long-established time frame. No longer. As the superintendent stated:

We looked at a natural curriculum picture before September, but we will address state priorities because our scores were awful. We weren't surprised; the student population we serve is the same as those at the bottom, the big city populations. We will try to raise scores in the third, fifth, and eighth grades. It doesn't mean they'll be smarter.

Another central office administrator detailed the changes more specifically:

We are building student anxiety, raising their level of concern. We don't want to do that with low esteem kids, so we're talking out of both sides of our mouths for our own political needs. Also, changes in math will be addressed in the normal math curriculum cycle next year, but this year we'll go ahead and make the changes in third, fifth, and eighth grades. Essentially, the [CSSO] just specified the third, fifth, and eighth grade reading and math curriculum. There is no local option because we have to spend more time on minimal curriculum than enrichment.

Once again, this district had relied on standardized tests in the past to gauge their instructional strengths and weaknesses. The assistant superintendent noted that,

In the past we've had more of a focus on [standardized tests]. Now the focus has shifted dramatically because we're looking for higher

scores in the third, fifth, and eighth grades on the state tests. They'll have more of an impact than the standardized test.

Clearly, administrators in these three districts were planning expedient strategies to improve the test scores, and just as clearly there was resentment to do so and a concern that what they were doing was compromising a standard of good professional practice. Essentially the message being given was that the test scores were becoming benchmarks for political reasons -- namely to appease school boards and communities who had had the opportunity to see their schools compared to one another and their system compared to neighboring districts, and who did not like what they saw. No matter how district staff had portrayed their performance in the past, part of that portrayal in the future had to include the test scores. Staff, in other words, were beginning to use the tests as a reference for judging local effectiveness.

This development reflected obligation more than acceptance. Perhaps most revealing was the ubiquitous "but" in their comments. Woven throughout the above passages were comments like "normally we do that, but now we have to do this." This syntactical form called attention to staff catching themselves in contradictions between what they publicly professed as good professional practice and what they found themselves actually doing. Put in terms of the dilemma Murnane (1987) stated, staff members were worried that specific attention to improving test scores would not improve learning.

Maryland

Maryland districts, subsequent to the RBS survey, seemed to be devoting more and more administrative and teacher time to devising strategies to improve scores on two of the tests and seemed increasingly to be using the scores as benchmarks, resulting in augmented pressure on teachers to get students to pass. Although no single event had dramatically heightened the stakes of the tests, students soon would have to pass all four of the tests in order to receive a diploma. The pressure to improve the percentage of students passing the tests increased dramatically following with each yearly test administration.

Not all four tests were regarded equally. Educators discriminated between the reading and math tests, on one hand, and the writing and citizenship ones on the other. The reading and math tests, in Maryland educators' minds, were adequate measures of basic competence in the respective content areas and covered objectives already well-entrenched in the curriculum. The curriculum development aspect of the state initiative began in the late seventies, and these two tests were the first to be developed, trial-tested, and implemented. Curriculum and instruction changes had been in place for seven to nine years in

some districts. By 1987, these adjustments had become institutionalized, to the point that interview subjects in four of the five districts argued that what was once novel was now routine.

> *We made sure everything we tested was in the curriculum. But that was done eight or nine years ago. The changes were already made [well before the survey]. [central office administrator]*

> *The [survey] mean [adjustments in curriculum and instruction] is skewed. Reading and math have been implemented for a while. [teacher]*

> *The changes in my area would have occurred well in the past. [teacher]*

The upshot was that the two tests were no longer obtrusive.

> *In reading, there probably hasn't been much change; the same in math. The scope and sequence were already complete and the content match was already there. [principal]*

> *Math and reading teachers probably don't have much of a problem anymore. [central office administrator]*

Such was not the case for the writing and citizenship tests. Both generated considerable controversy. The writing test did so because staff viewed it as demanding a performance level well beyond that necessary to be minimally competent in writing. The citizenship test's controversial aspect centered around its requirement that students memorize information about local, state, and federal governments -- information that even the teachers did not possess without special study. Fueling educators' concerns were the difficulties that a significant number of students were having in meeting the performance levels required by the two tests. Administrators, teachers with responsibilities in certain grades and in certain content areas, and special education teachers experienced growing pressure to improve the passing rate, adopting increasingly expedient methods of accomplishing this.

This "concentrated" approach to improving test results was apparent in all five districts, especially in schools where the scores were lowest. District 1 staff reported that considerable time was spent in preparation for the tests:

> *We are concentrating more on basics. We are now spending from September to November on basic skills rather than on our developmental program. [reading teacher]*

Another person complained that the writing test's importance was getting out of proportion:

> *The test has become the judge of the total system. [English teacher]*

Schools with low scores seemed to be getting special attention, as indicated in the following comment:

> *When the scores are low, [the poor performance] takes me into the school for the names of the kids who failed. There is no stroking in schools where scores have dropped. Everyone is sitting around with bated breath waiting for the test scores. [central office administrator]*

District 2 central office administrators agreed that the tests were assuming greater importance in the system, and the scores were a constant presence in their work.

> *Of course the tests are benchmarks. I always say it's only one indicator but it is the benchmark. It's reality. [central office administrator]*

> *The first question we ask is how we did relative to so and so. [central office administrator]*

> *Today I have 105 seniors who haven't passed. My anxiety is higher. [central office administrator]*

One administrator believed the pressure was greatest on schools with low scores.

> *I'm in the middle. I have no pressures at all. I know I'd feel uncomfortable on the bottom. [principal]*

District 3 seemed less consumed by the tests than other systems. Partly because of its small size, the burden of improving test performance fell on only a few shoulders. Moreover, the district had a history of deflecting the impact of state initiatives. Nevertheless, the tests had to be addressed.

> *We're bucking the system here. Many districts moved civics to the ninth grade and are testing for it in the tenth. We've had a program for a while in the twelfth grade. But it causes problems with no ninth grade civics class; we're interrupting classes to do a review. [teacher]*

> *I'm right now panically [sic] moving toward the test. [teacher]*

District 4 teachers were concerned about the extent to which passing the test was becoming an expediency in the district.

> *We realize a kid is taken out of science every other day for citizenship and will fail science to maybe pass the citizenship test. [building administrator]*

> *We're just getting them to memorize facts until [the test is given]. [teacher]*

> *I'm not opposed to the idea of testing, but I'm not sure we haven't gone overboard. The tail is wagging the dog. The original idea was that there were to be certain standards the student would have to meet, but if the student doesn't pass, people will ask what's wrong with the school and teachers. [teacher]*

These very targeted means for getting students to pass were acknowledged as a necessary evil:

> *We've had to do things we didn't want to do. [central office administrator]*

Staff in District 5 reported increasingly frequent interactions concerning how students were doing relative to the tests' objectives. They faced heightened awareness of the scores.

> *Teachers feel pressured to meet the superintendent's expected pass rate. [central office administrator]*

> *In administrators' meetings the talk is about where we rank. Parents let you know. You see it in newspapers. [principal]*

The result was the adoption of very focused strategies to teach test objectives in the classrooms.

> *Teachers feel jerked around. The test dictates what I will do in the classroom. [teacher]*

> *If you deviate from the objectives, you feel guilty, especially if kids fail. [teacher]*

> *We have materials provided by the county as 'quick help.' We were told 'here's how to get kids to pass the test fast.' They were good*

ideas but specifically on the test. For example, if the area in a rectangle is shaded, you multiply; if not, you add. [teacher]

And in response to the above stream of comments, a teacher summarized:

Talk about games and game-playing!

Reservations about strategies used to raise test scores were expressed in all five of the Maryland systems, just as they were in three of the Pennsylvania districts. As the importance of getting students to pass the tests heightened, local activity increasingly focused on the two troublesome tests, but in ways that produced the same linguistic qualifiers heard in Pennsylvania (most frequently "but"). Nevertheless, improving results became superordinate to other job responsibilities for many Maryland district administrators and a subset of teachers. Most of their professional time became devoted to test-related activities, to the exclusion of other staff development and improvement initiatives. This shift in job orientation seemed more widespread across the districts in Maryland than in Pennsylvania.

SHIFTING THE LOCAL FOCUS

It is important to note that the stakes -- the extent to which citizens and educators perceived that test performance would be used to make important decisions -- increased in the two states for two different reasons: (1) the SEA's use of the test scores to make comparisons of districts' performances in Pennsylvania and (2) the approach of the time when the results of all four tests would serve as an obstacle to graduation in Maryland. The stakes increased in what were originally both low and high-stakes situations. As they did, public pressure on districts to improve their performance intensified -- especially when a district's ability to improve seemed questionable (either because of the nature of the students or the nature of the test, or both), and/or when the need to demonstrate improvement was immediate (e.g, to correct unfavorable comparisons with other districts or between schools within a district). Educators' concerns shifted almost completely to influencing test performance. Put differently, a shift occurred in the seriousness with which the test was taken. The shift can best be described as a shift from a long-term to a short-term focus, from viewing the test as one indicator among many to treating the next set of test results as the most important outcome of schooling.

Such a shift is a probable occurrence in most rising stakes testing situations. In minimum competency testing, where the results are formulated typically in terms of the percentage of students passing the test, little technical expertise is needed to interpret the numbers. Thus, the results easily become

publicly accepted proxies for school performance. As the stakes associated with these readily intelligible numbers rise, the results also assume greater importance as statewide, standardized benchmarks -- and such benchmarks can become effective levers with which to move a district. Primarily because the public, and to an extent district staff members, hold the system as responsible for the performance of its students on the test, a need is created to gain control over activity that can influence those benchmarks. That is, the local community perceives the results as controllable, and the system undertakes an obligatory effort to do so. Moreover, students are the ones that directly suffer the consequences of failure in terms of being unable to graduate or move to the next grade, causing local educators to exert an even greater effort to improve student test performance. In the process, resources are drawn from other activities as staff members begin to analyze specific areas of student weakness on the tests and to develop materials directed specifically at improving performance. The more formidable the task of overcoming student weakness appears to be, and/or the more quickly improvement must be demonstrated, the more staff members devote their time to test-related activities.

Heightening this pressure to narrow the local focus is the cyclical nature of testing programs. The school year takes on a rhythmic quality with the tempo set by the test administration date. As the date approaches, activity directed toward improving performance becomes more frenetic. The test becomes foremost in the minds of the staff at least. The end result is that the major emphasis in the school becomes to improve the next set of scores rather than some longer-term, more general goal of improving student learning. Thus, the indicator of performance becomes the goal itself.

This recalls the dilemma stated by Murnane earlier: What if improving test scores does not improve student learning? Indeed, the key question in all of this discussion of stakes is, has learning improved or have only test results improved? The initial answer is that probably both occur. Focusing on improving the test scores of all students probably does lead to improved performance in general. But this works only up to a point. As the stakes rise and the pressure to perform better intensifies, activity becomes so focused on improving test scores that long-term learning opportunities are subordinated to efficient, short-term strategies to improve specific areas of weakness indicated by the test. Educators themselves verbally demonstrate the point at which this shift occurs through their use of linguistic qualifiers.

STAKES AND THE POLITICS OF EDUCATION

Perception. Pressure. Practice. This publication's message is that the perception of higher stakes associated with a state minimum competency test

leads to an intensification of the pressure on local educators to improve test scores, which in turn stimulates changes in local practice. Even though experts may regard some of these practices as appropriate (e.g., Popham et. al, 1985), the RBS research indicates that educators feel uncomfortable about the long-term value of many of their responses to high stakes testing. Improving the test results tends to become an end in itself, instigating considerable activity to improve the performance of "at-risk" populations through quick, intense preparation for the "day of the test."

Much of the pressure instigating these practices comes from the local community -- the newspaper, the school board, and parents. Actually these constituents seek to promote attainment of a desired level of an outcome rather than to encourage educators' engagement in specific practices. However, demanding particular levels of outcomes has been shown to be an especially effective means of exercising power over organizational action (Mintzberg, 1983). Power, according to Mechanic (1962:351), is "any force that results in behavior that would not have occurred if the force had not been present." Given the statements of the local educators detailed above, it is reasonable to assume they would not have engaged in many of the described practices in the absence of community pressure to improve test scores. Thus, outside influences became particularly potent factors in getting educators to behave in ways they ordinarily would not have.

In the specific instance with which this publication is concerned, knowledge of local performance on the test was the means of empowerment for various local constituencies. The test scores served as proxies for the quality of local educators' instructional behavior. In other words, how well teachers and administrators were discharging their educational responsibilities became visible through the window of test results. Increased visibility of one's performance improves the ability of others to reinforce behavior in accordance with expectations and to punish deviance (Merton, 1968; Nyberg, 1981). The information provided by the test enabled the community to determine whether its desired level of performance was being attained, and whether to attempt to influence district behavior.

The level of the stakes associated with mandated tests is the trigger for motivating external use of test scores as a lever to affect local practice. The community has other "objective" indicators upon which to base judgments about district performance and subsequent influence attempts. Whether an effort to shape district behavior ensues would seem to be related in large part to whether that indicator is used to make important decisions; the higher the stakes, the greater the pressure will be to correct performance deficiencies -- especially if improvement seems difficult or the need to demonstrate improvement is immediate.

High stakes statewide testing programs, then, can alter the political character of districts by increasing the probability that community elements can and will exercise influence. As Gutman (1988) explains, educators are accustomed to having to compromise the exercise of their professional judgment; citizen empowerment through their knowledge of test scores is just one of several barriers to the attainment of what she terms "appropriate levels of educators' autonomy" -- that is, autonomy that is neither so great as to shut out external influence altogether nor so insignificant as to make educators totally vulnerable to outside pressure. Johnson (1988) provides empirical evidence that teachers value highly this kind of "appropriate" autonomy and concludes that the key ingredient of current teacher reform proposals, if they are to produce better places for teachers to teach and students to learn, is the emphasis on enabling teachers to gain more control over their work. It seems, however, that the effects of high stakes testing on local control of education that were described above would countervail the most promising outcome of efforts to reform teaching.

Policy makers may want to consider ways to minimize one set of reforms negating another set. A significant step would be to lower the likelihood that scores alone will be perceived as the basis for someone's making important decisions. For example, if poor performance on the test triggered a district's engaging in a systematic, long-term improvement process rather than the denial of a symbol of progress like a promotion or a diploma, then the direct consequences for students would be lower -- as would the level of the stakes that the public probably would associate with the test. Likewise, creating alternative paths to graduation for seniors who fail a test (e.g., through teachers' and principals' certifying that students' demonstrated mastery of tested skills in homework or classwork) should accomplish much the same purpose. Doing the opposite, i.e., raising the stakes associated with a test, focuses attention solely on student performance and promotes the attainment of higher scores without necessarily improving learning. Such use of policy ultimately will undermine the very reforms it is supposed to encourage.

REFERENCES

Anderson, B., Odden, A., Farrar, E., Fuhrman, S., Davis, A., Huddle, E., Armstrong, J., & Flakus-Mosqueda, P. (1987). State Strategies to support local school improvement. **Knowledge: Creation, Diffusion, Utilization, 9**(1), 42-86.

Center for the Study of Testing, Evaluation, and Educational Policy (1986). An evaluation of the ICY essential skills tests in mathematics and reading. Chestnut Hill, MA: Boston College.

Corbett, H. D., & Wilson, B. L. (1987). Study of statewide mandatory minimum competency tests. Philadelphia, PA: Research for Better Schools.

First, J. M., & Cardenas, J. (1986). A minority view of testing. **Educational Measurement: Issues and Practice, 5**(1), 6-11.

Fuhrman, S. (1988). State politics and education reform. In R. Crowson & J. Hannaway (Eds.), **The politics of reforming school administration**. Philadelphia, PA: Falmer Press.

Gutman, A. (1988). Democractic theory and the role of teachers in democratic education. In R. Crowson & J. Hannaway (Eds.), **The politics of reforming school administration**. Philadelphia, PA: Falmer Press.

Johnson, S. (1988). Schoolwork and its reform. In R. Crowson & J. Hannaway (Eds.), **The politics of reforming school administration**. Philadelphia, PA: Falmer Press.

LeMahieu. P. G. (1984). The effects on achievement and instructional content of a program of student monitoring through frequent testing. **Educational Evaluation and Policy Analysis, 6**(2), 175-187.

Madaus, G. (1988). The influence of testing on the curriculum. In L. Tanner (ed.), **Critical issues in curriculum: 87th Yearbook of the NSSE, Part I**. Chicago, IL: University of Chicago Press.

McDonnell, L., & Elmore, R. (1987). Getting the job done: Alternative policy instruments. **Educational Evaluation and Policy Analysis, 9**(2), 133-152.

Mechanic, D. (1962). Sources of power of lower participants in complex organizations. **Administrative Science Quarterly**, 349-364.

Merton, R. (1968). **Social theory and social structure**. Chicago, IL: Free Press.

Mintzberg, H. (1983). **Structure in five: Designing effective organizations.** Englewood Cliffs, NJ: Prentice Hall.

Murnane, R. (1987). Improving education indicators and economic indicators: The same problem? **Educational Evaluation and Policy Analysis, 9**(2), 101-116.

Nyberg, D. (1981). A concept of power for education. **Teachers College Record, 82**(4), 535-551.

Polemini, A. J. (1977). Security in a citywide testing program. **Journal of Teaching and Learning, 2**(3), 34-40.

Popham, W. J., Cruse, K. L., Ranking, S. C., Sandifer, P. W., & Williams, P. L.,(1985). Measurement-driven instruction: It's on the road. **Phi Delta Kappan, 66**(9), 628-634.

Stake, R., Bettridge, J., Metzer, D., & Switzer, D. (1987). Review of literature on effects of achievement testing. Champaign, IL: Center for Instructional Research and Curriculum Evaluation.

Research for Better Schools (RBS), a private, non-profit, educational research and development firm, was founded in 1966. Its sponsors include many clients from the public and private sector who support R&D projects that meet their needs. RBS is funded by the U.S. Department of Education to serve as the educational laboratory for the Mid-Atlantic region.

Using the expertise of some 50 staff members, RBS conducts research and policy studies on key education issues, develops improvement approaches and services for schools, provides consultant services to state leaders, develops products for special populations, and participates in national networking activities with other regional laboratories to enhance the use of R&D products and knowledge.

During the past 20 years, RBS has developed extensive capabilities which are available to all education professionals in the form of practical, research-based products and services. This publication is one of the products of RBS' R&D work. Related training and technical assistance services also are available. Your interest in RBS is appreciated and your suggestions or requests for information always are welcome.